DOCUMENT RESUME

ED 267 523 EC 182 042

AUTHOR Fujiki, Martin; Brinton, Bonnie
TITLE Sampling Reliability in Spontaneous Language Sampling.
PUB DATE Nov 85
NOTE 13p.; Paper presented at the Annual Convention of the American Speech-Language-Hearing Association (12th, Washington, DC, November 22-25, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Language Handicaps; *Language Tests; *Sampling; *Syntax; Test Reliability; Young Children

ABSTRACT

To determine how many occurrences of a syntactic structure are necessary to provide sampling reliability, two one-half hour spontaneous language samples were elicited from each of 15 language disordered students (5 to 6 years old). Sessions were divided into two periods, one for the child's telling about pictures and toys and the other for conversation between examiner and child. Thirteen syntactical structures were examined for analysis. Proportion of occurrence for each structure was calculated, a difference score (difference between the total number of occurrences and sampling levels) computed, and a mean absolute error proportion across all syntactic structures for each sampling level derived. Data suggested that a set of four trials was a usable guideline for sampling reliability in spontaneous language production, so that if a child produces a structure on at least four unique occasions, the clinician may assume that this provides a representative picture of the child's knowledge of that structure. (CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED267523

Sampling Reliability in Spontaneous Language Sampling

Martin Fujiki & Bonnie Brinton

Paper presented at the American Speech-Language-Hearing Convention, Washington, D.C., 1985

Spontaneous language sampling is generally considered to be the method of choice in the evaluation of language-disordered children. It is widely used by speech-language pathologists to both separate normal from language-disordered children and to provide information concerning treatment goals. Although it is generally assumed that this procedure provides a reliable measure of the child's linguistic system, there are few studies available examining the reliability of this procedure. One specific aspect of reliability that has clinical significance is the question of how many occurrences of a structure are necessary to achieve sampling reliability. For example, it might reasonably be assumed that a single occurrence of a structure does not constitute a sufficient data base from which clinical decisions might be made. However, the question of how many times a structure must appear to provide stable data has not been experimentally examined.

This question has been indirectly considered with regard to structure productivity and acquisition. For example, Morehead and Ingram (1973) studied the syntactic development of normal and linguistically deviant children. In doing so, these authors considered two or more appropriate occurrences of a structure to represent rule productivity.

EC182042

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Martin Fujiki

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Bloom (1970) set a criterion for structure productivity as five or more appropriate productions of the structure in the course of five to eight hours of observation. Bloom and Lahey (1978) later modified this criterion for clinicial application, suggesting that four unique productions in an hour session would indicate productivity.

However, as Lahey (1983) has pointed out, productivity does not necessarily indicate acquisition. With regard to structure acquisition, Brown (1973) noted that there may be a considerable time lapse between the initial appearance and mastery of a form. Because of this, Brown proposed that the percentage of appropriate use in obligatory context be examined to establish acquisition. Using this method, language samples were examined for contexts requiring the use of the structure in the adult rule system. The first of three consecutive samples in which the child used the structure at the 90% appropriate level in obligatory contexts was considered as the point of acquisition.

Although procedures similar to Brown's methodology have been successfully employed by a number of researchers (deVilliers and deVilliers, 1973; Ingram, 1974; Johnston and Schery, 1976) this method only indirectly addresses the question at hand. Further, although the researcher who has access to a large corpus of language data may obtain an accurate picture of the child's abilities in this manner, the typical clinician is often limited by the amount of language data required to make such determinations. The clinician, out of necessity, must be concerned not only with obtaining an accurate picture of the

child's knowledge, but also with doing so as efficiently as possible. The purpose of this study was to determine how many occurrences of a syntactic structure are necessary to provide sampling reliability. Data addressing this question would allow the clinician to more efficiently use spontaneous language sampling by providing an indication of how many occurrences of a structure are necessary to provide a stable picture of a child's knowledge concerning that structure. In that this question is of most importance in a clinical sense, the study examined the performance of language-disordered children.

METHOD

15 subjects were selected from children with diagnosed language disorders from two elementary schools in Sandy, Utah. All of the subjects were sampled from the age range of 5:6 to 6:6 and were from monolingual home environments. The elementary schools involved in the study were located in neighborhoods with similar property values. The following criteria were used to define the term "language-disordered."

1. All subjects were diagnosed as having a language disorder by the school speech pathologist, and were enrolled in language therapy.

2. On the basis of spontaneous language sampling, all subjects had not mastered at least four common, age appropriate grammatical morphemes. Mastery criterion was set at 90% appropriate usage in obligatory context, with a minimum of four unique occurrences.

3. All subjects demonstrated a score below the 15th

percentile on a formal measure of language development. Formal measures used to provide these scores included the Carrow Elicited Language Inventory (Carrow, 1974) and the Bankson Language Screening Test (Bankson, 1977).

4. All subjects demonstrated normal physical, emotional, and intellectual development on the basis of school records and observation by the classroom teacher and the school speech pathologist.

Two one-half hour spontaneous language samples, providing a mean of 449 utterances (range from 299 to 654), were elicited from each subject. Testing was performed in two sessions, with all sessions for an individual subject being performed within one week's time.

Each language sample was divided into two general periods of continuous tape recording. In one 15 minute period the child was provided with toys and pictures. Picture stimuli consisted of illustrations from the Motivation and Learning Centered Training Programs for Language Delayed Children (Mecham, 1974). Subjects were asked to tell about the pictures, which represented specific actions desgined to elicit structures such as the auxiliary be forms. Toy stimuli included a Fisher-Price School and other similar materials. Subjects were allowed to engage in free play without examiner direction. The other 15 minute period consisted of conversation between examiner and child. Procedures based upon methods suggested by Labov (1972), including verbal activities such as describing how to play a game, how to perform a particular activity, and telling about an emotional event, were used to elicit language production. The order of the 15 minute segments was systematically varied to control order effects. On occasion a period was extended beyond 15 minutes if it

appeared to be especially well suited for a specific child. Thus, not all of the samples were divided into exact 15 minute halves. However, this variability was generally small. All of the language samples were tape recorded using a Sony TC-161 SD cassette-recorder with Maxell low-noise cassette tape.

Test Reliability

Naturalistic language sampling transcription reliability was evaluated as follows. Two randomly selected taped samples were transcribed by a second examiner to obtain an estimate of reliability. The percentage of agreement between the examiners regarding transcription accuracy was calculated using the following formula: [B + C/N] x 100, where B=the number of agreements of presence of a structure, C=the number of agreements of absences, and N=the total number of observations. This produced a percentage of agreement of 96%. Scoring reliability was calculated using the same formula, and a percentage of agreement of 94% was obtained.

Analysis of the Data

The following 13 syntactical structures were examined for analysis: auxiliary is, auxiliary are, copula is, copula are, regular past ed, regular plural s, prepositions in and on, pronouns I, he, she, and articles a and the. In order to determine sampling reliability, the

following method was used (Brookshire, Nicholas, & Krueger, 1978; Fujiki & Brinton, 1983). The proportion of occurrence (appropriate occurrence: total occurrence) for each structure tested was calculated for the data produced by the following sampling levels: the first occurrence, the first three occurrences, the first five occurrences, the first seven occurrences, and 10 occurrences. On the basis of 10 occurrences, the 80% confidence interval was calculated for each of the syntactic forms.

In addition, for each structure the proportion occurrence generated by each sampling level was subtracted from the total proportional occurrence (the proportion of occurrence based on 10 occurrences of the structure). This produced a difference score (the difference between the total number of occurrences and the sampling levels) for each of the syntactic structures. These scores were then summed ignoring minus signs to avoid cancellation of positive sampling error in one category by negative sampling error in another. This total was divided by 13 (the number of structures tested) to produce a mean absolute error proportion across all of the syntactic structures for each sampling level. To be considered as a usable sampling level, the error proportion of a particular level was required to be within 10% of the total proportional occurrence, when both mean and standard deviation were taken into consideration.

RESULTS

Table 1 presents the mean percentages of appropriate production in spontaneous language sampling for the first, the first three, the first five, the first seven, and the first 10 productions in spontaneous language sampling. The 80% confidence interval is also presented.

Because of the nature of spontaneous language production, it was not possible to control the frequency of occurrence of the structures under study. Thus, it was necessary to exclude subjects who failed to produce a structure on ten or more trials from the analysis of that structure. The number of subjects included in the analysis of each structure is also reported.

insert Table 1

Of the 13 structures studied, seven produced means on the first sampling level that fell within the 80% confidence interval of the mean of 10 trials. One of these forms (auxiliary _are_) produced a mean outside of the confidence interval on the three and five trial levels. Of the remaining structures, one produced a mean within the confidence interval of 10 trials on the three trial level, and three produced a mean within the confidence interval on the five trial level. Only one structure required seven occurrences to produce a mean within the 10 trial level. This finding suggested that five trials was the best guide for sampling reliability in this particular situation. Such a

guideline would have produced inaccurate results for only one structure.

Table 2 presents the mean percentage differences and standard deviations from 10 occurrences for the 1, 3, 5, and 7 occurrence sampling levels.

insert Table 2

When standard deviation was considered, the mean for the one and three trial levels fell outside of the 10% range previously established as defining a usable sampling level. This placed the cut off point between the three and five trials levels, and raised the question as to whether or not four trials might be a viable sampling level. Therefore, the four trial level was also examined. It was found that the four trial level produced a mean proportion difference of 4.50 and a standard deviation of 4.52. Thus, the four trial level was also within the previously established 10% range.

DISCUSSION

On the basis of these data, it appeared that four trials was a usable guideline for sampling reliability in spontaneous language production. It was of interest that acquired structures (based on a group performance mean above 90%), while being produced at high levels of

appropriate usage with little variability, did not consistently produce stable results at lower sampling levels than unacquired forms. However, it was observed that the two structures that were produced at the lowest levels of appropriate performance (auxiliary is and are) required greater numbers of trials to produce sampling reliability (five and seven trials respectively).

Regarding clinical application, these findings indicate that the four, five, and seven trial sampling levels produced results that were reliable with results produced by the 10 trial level. Thus, at least in the current study, a child's performance on four or more trials was representative of his performance on 10 trials. This suggests that if a child has produced a structure on at least four unique occasions, the clinician may assume that this provides a representative picture of the child's knowledge of that structure. It is of interest that this finding provides indirect support for Bloom and Lahey's (1978) standard of four unique productions being an indication of productivity.

Considering the variability inherent in naturalistic language production, these data must be interpreted cautiously. Further more, it is recognized that these results were obtained from a limited number of subjects being sampled in a controlled situation. Manipulation of any of these variables might result in differing results. However, these data do present a general guide by which sampling reliability can be assessed. Additional research, performed with larger samples and examining a wider range of forms, may serve to further clarify this issue.

REFERENCES

Bankson, N. (1977). Bankson Language Screening Test. Baltimore: University Park Press.

Bloom, L. (1970). Language development: Form and function in emerging grammars. Cambridge, MA: MIT Press.

Bloom, L. & Lahey, M. (1978). Language development and language disorders. New York: John Wiley & Sons.

Brookshire, R.H., Nicholas, L.S., & Krueger, K. (1978). An evaluation of momentary and interval sampling procedures. Journal of Speech and Hearing Research, 21, 652-667.

Brown, R. (1973). A first language. Cambridge, MA: Harvard University Press.

Carrow, E. (1974). Carrow elicited language inventory. Austin, Texas: Learning Concepts.

deVilliers, J.G. & deVilliers, P.A. (1973). A cross-sectional study of the acquisition of grammatical morphemes. Journal of Psycholinguistic Research, 2, 267-278.

Fujiki, M. & Brinton, B. (1983). Sampling reliability in elicited

imitation. Journal of Speech and Hearing Disorders, 48, 85-89.

Johnston, J.R. & Schery, T.K. (1976). The use of grammatical morphemes by children with communication disorders. In D.M. Morehead & A. E. Morehead (Eds.) Normal and deficient child language. Baltimore: University Park Press.

Ingram, D. (1974). The acquisition of the English verbal auxiliary and copula in normal and linguistically deviant children. In L. McReynolds (Ed.), Developing systematic procedures for training children's language, ASHA Monographs, 18, 1974.

Labov, W. (1972). Sociolinguistic patterns. Philadelphia: University of Pennsylvania.

Lahey, M. (1983). After the simple sentence: Assessing and facilitating complex content/form/use interactions. Jordan School District Workshop, Sandy, Utah.

Mecham, M. (1974). Motivation and Learning Centered Training Programs for language Delayed Children. Salt Lake City, Utah: Word Making Productions.

Morehead, D.M. & Ingram, D. (1973). The development of base syntax in normal and linguistically deviant children. Journal of Speech and Hearing Research, 16, 330-352.

Table 1. Mean proportions of appropriate production in spontaneous sampling for the 1, 3, 5, 7 and 10 trial sampling levels, and the 80% confidence interval based on 10 trials.

Syntactic Structure		Subjects	Sampling Level					80% CI for 10 trials
			1	3	5	7	10	
1.	auxiliary is	14	43	45	57	60	60%	47-71%
2.	auxiliary are	5	40	33	36	43	48%	37-53%
3.	copula is	14	86	83	83	84	80%	73-87%
4.	copula are	5	60	47	44	51	48%	33-63%
5.	reg. past	6	83	78	70	69	67%	57-77%
6.	reg. plural	15	100	98	93	90	90%	85-95%
7.	in	11	91	94	95	96	96%	93-99%
8.	on	12	100	94	93	94	94%	91-97%
9.	I	15	100	98	99	98	97%	95-99%
10.	he	12	75	75	73	74	74%	60-89%
11.	she	7	57	62	60	61	67%	48-86%
12.	a	15	87	82	81	81	83%	77-89%
13.	the	15	89	85	85	87	85%	79-91%

note. number of subjects indicates the number producing the structure on 10 or more trials.

Table 2. Mean proportion difference and standard deviation from 10 occurrences for the 1, 3, 5, and 7 occurrence sampling levels.

	Trials			
	1	3	5	7
Mean (%)	7.85	4.77	3.23	1.92
SD	4.74	5.35	3.05	1.98